

THE UNIVERSITY OF BURDWAN



TOPIC : ACADEMIC PERFORMANCE ANALYSIS

Name: JAYEETA PAL

Roll No: BUR/ST/2021/007

Reg. No: 202103000230 of 2021-22

Subject: Statistics

Semester: 3rd Semester

Course: Community Engagement Programme

Session: 2022-2023

Statistical Analysis of the factors affecting students' academic performance in Post - Graduation : A case study of the department of statistics, The University of Burdwan, West Bengal

Jayeeta Pal, student of the department of statistics, The University of Burdwan

Abstract: This research examined factors affecting first year students' academic performance in post-graduation taking the department of statistics, The University of Burdwan, West Bengal as a case study. Seven factors namely: sex, family income, study time of a student per day, social media involvement per day, B.M.I of a student, B.Sc. marks, favorite topics in between theoretical statistics and applied statistics, were considered. A sample of 22 students of first year students from the department of statistics was selected purposively and data was obtained through a structured questionnaire with a response value of 22 students. Multiple regression analysis revealed a model which proved to be significant at 0.05 level of significance by analysis of variance and a multiple correlation value of 0.9426 which indicates a strong positive relationship between SGPA and the

predictors. The coefficient of determination posited that on and above 90% of the variation in academic performance was explained by the predictor variables. The result of the analysis indicated that study time per day, social media involvement per day, b.sc. marks, family income contribute to students' academic performance(SGPA).

Keywords: Academic performance, multiple regression, SGPA, factors, post-graduation, students

1. Introduction:

Students' academic gain and learning performance is affected by numerous factors including b.sc. marks , parent/guardian socio-economic status, daily study hour etc.

Academic Performance SGPA

Academic performance refers to the level of performance in school, accomplishment or success in school. However, academic performance is the core of educational growth.

Statement of the Problem

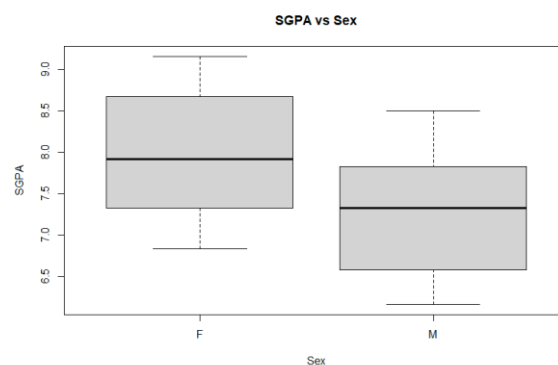
The most prevalent argument is that the socioeconomic status of learners affects the quality of their academic performance. Student academic performance measurement has been a challenging issue in social, educational, psychological, environmental and personal factors. The utility of these studies lies in the need to undertake corrective measures that improve the academic performance of post-

graduate students. Among all factors, socioeconomic status is one of the most researched and debated factor among educational professionals that contribute towards the academic performance of students. These factors strongly influence students' academic pursuit though they vary from person to person. This study, therefore aims to identify and analyse some determinant factors of academic performance in post-graduate students. This research will therefore consider family socio-economic status, sex, study hour per day, social media using time per day etc. that determine the academic performance of the students.

Now we see the effects of the categorical variables- family income, sex, social media involvement per day, favorite topics (theoretical statistics or applied statistics) over the academic performance of the first year post-graduation students of the department of statistics of the university.

i) Relationship between sex and SGPA

To examine the relationship between sex and SGPA, first we saw a boxplot of these two factors. Here the following diagram is below:



From this boxplot, we see female students get greater marks in examinations than the male students of the university. So the academic performance (SGPA) for female students is higher than the male students. The distribution of the data for male students is slight positively skewed and the distribution of the data for female students is more or less symmetric.

Then we applied the analysis of variance (ANOVA) technique to test for the equality of population mean for the two categories of the students - male and female.

R Code:

```
# Boxplot of SGPA vs Sex
```

```
boxplot(c1$cgpa~c1$Sex,main='SGPA vs Sex',xlab="Sex",ylab="SGPA")
```

```
# Analysis of Variance Model for SGPA vs Sex
```

```
anova1<-
```

```
aov(c1$cgpa~c1$Sex,data=c1)
```

```
anova1
```

```
summary(anova1)
```

Output:

```
> summary(anova1)
```

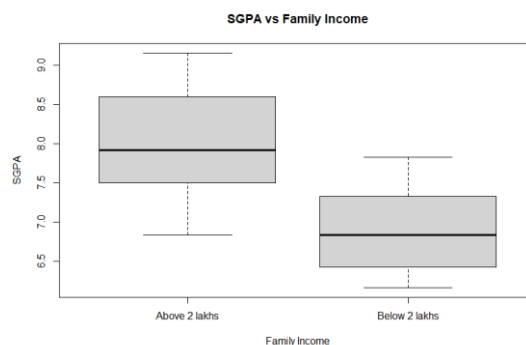
| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|----------|
| C\$Sex | 1 | 2.776 | 2.776 | 4.378 | 0.0494 * |
| Residuals | 20 | 12.680 | 0.634 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Here, we can see that p-value is 0.0494 which is slight less than 0.05, level of significance. So, we can conclude that null hypothesis of equal population mean for the two categories- male and female students, is rejected with 95% confidence level. So, the population mean of SGPA of female students is different from the population mean of SGPA of male students of the statistics department of the above university and from the boxplot we saw that SGPA of female students is higher than the SGPA of male students. So, we can conclude that academic performance (SGPA) of the students is highly affected by their sex.

ii) Relationship between Family Income and SGPA

Considine and Zappala (2002) also noticed that parent's income or socio-economic condition positively affects the student test score in examination. Here, to examine the relationship between family income and SGPA, first we saw a boxplot of these two factors. Here the following diagram is below:



From this boxplot, we see the students who belong from a comparatively rich families get greater marks in examinations than the students of poor family of the university. Here we divided the socio-economic condition of the students in two categories- a. family income is 2 lakhs and above, b. family income is below 2 lakhs. So the academic performance (SGPA) of first category is higher than the students belong to the second category. The distribution of the data for first category students is slight negatively skewed and the distribution of the data for second category students is slight positively skewed.

Then we applied the analysis of variance (ANOVA) technique to test for the equality of population mean for the two categories of the students according to their family income.

R Code:

```
# Boxplot of SGPA vs Family Income
```

```
boxplot(c1$cgpa~c1$`Family income per annum`,main='SGPA vs Family Income',xlab="Family Income",ylab="SGPA")
```

```
# Analysis of variance Model for SGPA vs Family Income
```

```
anova2<-aov(c1$cgpa~c1$`Family income per annum`,data=c1)
```

```
anova2
```

```
summary(anova2)
```

Output:

```
summary(anova2)
```

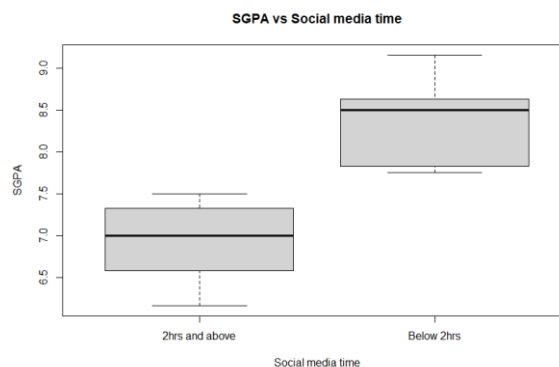
| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-------------------------------|----|--------|---------|---------|--------------|
| c1\$`Family income per annum` | 1 | 6.901 | 6.901 | 16.13 | 0.000676 *** |
| Residuals | 20 | 8.554 | 0.428 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Here, we can see that p-value is 0.000676 which is very lower than 0.05, level of significance. So, we can conclude that null hypothesis of equal population mean for the two categories according to family income of the students is rejected with 95% confidence level. So, the population mean of SGPA of students belong to the families of family income 2 lakhs and above is different from the population mean of SGPA of students belong to the families of family income below 2 lakhs among the students of the statistics department of the above university and from the boxplot we saw that SGPA of students of first category is higher than the SGPA of students belong to the second category. So, we can conclude that academic performance (SGPA) of the students is highly affected by their socio-economic condition of their families.

iii) Relationship between Social media involvement time and SGPA

To examine the relationship between social media involvement per day and SGPA, first we saw a boxplot of these two factors. Here the following diagram is below:



From the above boxplot, we see the students who waste their time more in social media, get lower marks in examinations than the students who invest comparatively less time in social media per day. Here we divided our data according to social media activeness in two categories- a) below 2 hours and b) 2 hours and above. So the academic performance (SGPA) of first category is higher than the students belong to the second category. The distribution of the data for first category students is more or less symmetric and the distribution of the data for second category students is highly positively skewed.

Then we perform the analysis of variance technique or ANOVA technique to test for the equality of population mean

for the two categories of the students according to their social media involvement time per day of the students.

R Code:

```
# Boxplot of SGPA vs Social media time

boxplot(c1$cgpa~c1$`How much time you spend in social media per
day?`,main='SGPA vs Social media time',xlab="Social media
time",ylab="SGPA")

# Analysis of Variance Model for SGPA vs Social Media Time

anova3<-aov(c1$cgpa~c1$`How much time you spend in social media per
day?`,data=c1)

anova3

summary(anova3)
```

Output:

```
summary(anova3)
```

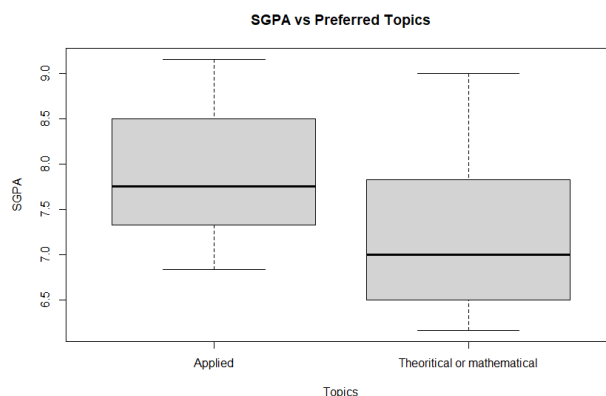
| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|----|--------|---------|---------|--------------|
| c1\$`How much time you spend in social media per day?` | 1 | 10.710 | 10.710 | 45.14 | 1.55e-06 *** |
| Residuals | 20 | 4.745 | 0.237 | | |
| --- | | | | | |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | | |

Here, we can see that the p-value is very lower than the level of significance 0.05. So, we can conclude that null hypothesis of equal population mean for the two categories according to social media using time of the students is rejected with 95%

confidence level. So, the population mean of SGPA of the students of first category (below 2 hours) is different from the population mean of SGPA of the students of second category (2 hours and above). Besides, from the above boxplot we can see that the SGPA of the students of first category is higher than the second category. Therefore, we can conclude that academic performance (SGPA) of the students is highly negatively affected by their involvement period in social media per day.

iv) Relationship between Favorite Topics and SGPA

To check the relationship between favorite topics and SGPA, first we draw a boxplot of these two factors. Here the following diagram is below:



From the above boxplot, we see the students who preferred theoretical and mathematical topics of statistics get lower marks in examination than the students whose favorite topic is applied statistics. Here we divided our data according to the students preferred topic- a) Theoretical and

mathematical topics and b) applied statistics. So, the academic performance (SGPA) of first category is lower than the students belong to the second category. The distribution of the data for first category students is slight negatively skewed and the distribution of the data for second category students is also negatively skewed.

Then we perform the analysis of variance technique or ANOVA technique to test for the equality of population mean for the two categories of the students according to their preferred topics.

R Code:

```
# Boxplot of SGPA vs Preferred Topics
```

```
boxplot(C$cgpa~C$`Preffered Topics`,main='SGPA vs Preferred  
Topics',xlab="Topics",ylab="SGPA")
```

```
# Analysis of Variance Model for SGPA vs Preferred Topics
```

```
anova4<-aov(c1$cgpa~c1$`Preffered Topics`,data=c1)
```

```
anova4
```

```
summary(anova4)
```

Output:

```
summary(anova4)
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|--|----|--------|---------|---------|--------|
|--|----|--------|---------|---------|--------|

| | | | | | |
|------------------------|---|-------|--------|-------|-------|
| c1\$`Preffered Topics` | 1 | 1.574 | 1.5739 | 2.268 | 0.148 |
|------------------------|---|-------|--------|-------|-------|

| | | | | | |
|-----------|----|--------|--------|--|--|
| Residuals | 20 | 13.882 | 0.6941 | | |
|-----------|----|--------|--------|--|--|

But from the output of ANOVA technique we see the p-value is more than the level of significance 0.05. So, we can conclude that null hypothesis of equal population mean for the two categories according to preferred topics of the students is accepted with 95% confidence level. Therefore, the population mean of SGPA of the students of first category who like theoretical and mathematical part of statistics is almost same with the population mean of the second category students who preferred applied topics in statistics more. But, from the boxplot we see the students who choose applied topics, get more marks in examination. So, here we can't conclude confidently that the preferred topics highly affect students' academic performance.

